



APPENDIX III

Machines for Searching Recorded Information: IBM Electronic

Scanning System, Mark III

I. Introduction

At present (April 1953) it appears that no machine now in production is fully suitable for searching large quantities of recorded information effectively and practically. In recognition of the need for such machines the International Business Machines Corporation, in its Engineering Laboratory at Poughkeepsie, New York, has given considerable thought to possible designs. A model of one design (IBM Electronic Scanning System, Mark I) was built in 1950 and is now in use for experimental purposes. An improved design (Mark II) was later developed but not built. Additional improvements in effectiveness and practicality have been incorporated recently in a third design (Mark III). Punched IBM cards provide the searching medium in all three designs. A note-worthy feature of the latest design (Mark III) is that it integrates the searching system with operations based on standard IBM equipment: The searching machine, a scanner, may be complemented by standard rather than special auxiliary equipment, and provision is made for the searching of files of IBM cards prepared in the conventional manner for processing on the basis of fixed fields. A description of the characteristics and capabilities of the Mark III system is provided below. A model of this equipment has not yet been built, and the terms of its procurement have not yet been requested.

II. Description of the IBM Electronic Scanning System, Mark III

a. Recording System

1. Recording Medium

Conventional IBM cards (80 columns by 12 rows) will be used as media for recording information in a mechanically searchable form.

2. Recording Technique

- (a) The basic operation of recording will consist in key-punching a pattern of one or more holes into

a single column of a card. Any IBM key punch may be used for this purpose.

(b) Expressions involving the use of more than one column will be recorded by key-punching the appropriate patterns into successive columns on a card.

(c) As in writing or typing, a desired expression (for example, a proper name) may be recorded anywhere on a card, and the same expression may be recorded in different locations on different cards. The method of searching will be such that recognition of a pattern or succession of patterns is not dependent upon field location. By eliminating the necessity to record in fixed fields the system achieves the following features:

(1) It will not be necessary to limit arbitrarily the length, variety, or number of recorded expressions.

(2) It will not be necessary to sacrifice card capacity by reserving space for expressions which are not needed to represent the information at hand.

(3) It will be possible to introduce new expressions at will, as needed.

(It is assumed that an identifying number or other means of identification will be recorded at the beginning of every record card, to implement operations of sorting and control.)

(d) All of the eighty (80) columns on a card may be used for recording mechanically searchable information. If the information recorded on a card does not require the punching of all eighty columns, the remaining columns may be left unpunched. (The "lacing" of unused columns employed in the Mark I design will not be required.)

3. Amount of Information Recorded per Source

The system will be flexible with respect to the length of record punched for a single source (for example, a single document). As many cards as needed may be used to represent a source. (See section g for further details.)

~~CONFIDENTIAL~~
Security Information

4. Punching Codes

The system will be flexible with respect to the type of punching code employed.

(a) Standard IBM Code

Standard IBM code may be used directly for recording information. Special applications of this code (for example, multiple coding in a single column) also may be utilized. When standard IBM code is used for recording in the Mark III system, no special equipment is needed other than the searching machine itself; punching, verifying, etc. will be accomplished by means of standard IBM equipment.

Standard IBM code consists of an assignment of forty-seven (47) columnar punch-patterns to forty-seven printed characters (see Figure I, p. 10, for details). The following groups of characters are provided:

<u>Characters</u>	<u>Number of Holes Punched per Column</u>
10 numeric (0-9)	1
26 Alphabetic (A-Z)	2
11 special (punctuation, etc.)	1, 2, or 3

(b) The IBM Scanning Code

The IBM scanning code employed in the Mark I system may be utilized also in the Mark III system. For a practical use of this code special auxiliary equipment (punch, transcriber, etc.) is needed, in addition to the searching machine itself.

The IBM scanning code is based on an "alphabet" of 792 punch-patterns. Each of the 792 patterns consists of a combination of five holes in a twelve-place column.

(c) Other Punching Codes

Standard IBM code and the IBM scanning code may be regarded as basic "alphabets" of the Mark III system, since special provisions will be made for searching these codes conveniently. It will be possible, however, to record and to search any combination of holes that can occur in a twelve-place column.

b. Method of Searching

The mechanization of searching will be based upon the scanning

Security Information

and matching method. The eighty columns or pattern-positions of a record card will be examined consecutively, one by one. A single inspection of each column will suffice to determine the presence or absence of the pattern sought. A pattern or succession of patterns will be recognized as such, without respect to field location. (Field Location can be taken into account in searching, however, if meaning has been assigned to it; see section h.)

c. Speed of Searching and Capacity of Card Input Hopper

The rate of searching will be in the range of 600 to 1000 cards (48,000 to 80,000 columns or pattern-positions) per minute. Cards will be fed to the scanner in lots containing as many as 2000 cards per lot.

d. Mechanism of Searching

1. Identification

The identification of those sources which meet the requirements of a search will be accomplished by the scanner as follows:

(a) Punched information will be read from moving record cards photoelectrically and will be transmitted to a matching component.

(b) In the matching component the received information will be compared with the requirements of the search by electronic means.

(c) When any one of the requirements of a search is met, the fact that it has been met will be "stored" electromagnetically (i.e., in relays). Such storage will persist until a clearing condition is encountered (see section e3 for further details).

2. Selection

When a given source has been identified as one which meets in full the requirements of the search, the cards representing that source will be selected automatically. Cards representing sources which fail to meet all of the requirements of the search will be "rejected". (In lieu of or in addition to physical selection, counting operations will be possible; see section e4 for details.)

Security Information

e. Scope and Flexibility of Searching Operations

1. Searching Capacity

(a) Number and Length of Expressions

A single search may be directed to several expressions, and these expressions may vary in length. (An 'expression' is any columnar pattern or succession of columnar patterns which functions as a unit for recording or searching - for example, a word or phrase.)

(1) Number of Expressions

The number of expressions that may be sought on a single pass of cards will be variable from one (1) expression to an initial maximum of twelve (12) expressions*. (The maximum of twelve expressions is exclusive of requirements for the execution of clearing functions described in section e3.)

(2) Length of Expressions

The length of each sought expression will be variable from one (1) column (as in the case of a one-letter word) to any desired number of columns, subject to the overall requirement that the total length for all expressions be eighty (80) columns * or less for a single pass of cards.

2. Ways of Combining Expressions

The system will be flexible with respect to ways in which the component expressions of a search may be combined. The following types of combination or conditioning will be available.

*The indicated maximum can be increased by relatively minor changes in the equipment, if need for such increase develops.

Security Information

(a) Combinations of Two Expressions

Any two expressions (A, B) (for example: 'house', 'cat') may be combined in any one of the following ways:

<u>Combining Operator</u>	<u>Combining Conditions: Execute required function if and only if, within the boundaries specified</u>
(1) conjunction	(A is present) and (B is present)
(2) disjunction	(A is present) and/or (B is present)
(3) negation-conjunction	(A is present) and not-(B is present)
(4) sequenced conjunction	(A is present) and (B is present) and (A precedes B)
(5) tied conjunction**	(A is present) and (B is present) and (A precedes B) and (A adjoins B)

(b) Combinations of three or More Expressions

Any three or more expressions (A, B, C, ---) may be combined by repetitive application of one combining operator or by a combination involving two or more combining operators.

Examples

Repetitive Conjunction	(A is present) and (B is present) and (C is present)
Repetitive Disjunction	(A is present) and/or (B is present) and/or (C is present)
Mixed Combinations	((A is present) and (B is present)) and/or ((C is present) and (D is present)) ((A is present) and/or (B is present)) and not-(C is present)

3. Control of Boundaries: Clearing Functions

It will be possible to control the boundaries within which a combination of recorded expressions is required to occur. This will be accomplished by the provision

of clearing functions, the effect of which will be to "erase" stored information if, at the boundary specified, the stored information does not fully satisfy the corresponding search requirement. In all normal searching operations, exhaustive clearing ("erasing") will be effected when the end (or beginning) of the record for a given source is reached. In addition to provision for this general type of clearing, five (5) subordinate levels of clearing will be available for use when needed.

- (a) Any desired pattern or succession of patterns may be selected to serve as a clearing expression.
- (b) In a given search it will be possible to use more than one clearing expression and to use a particular clearing expression more than once.
- (c) Execution of subordinate clearing functions will be optional; if the requirements of a given search do not involve subordinate clearing, the presence of clearing expressions in the passing record cards may be ignored.
- (d) Provision for the inclusion of clearing expressions in a search is in addition to the provision for as many as twelve expressions which do not have this function.

4. Output Functions

Two output functions will be provided - card selection and unit counting. For card selection, one selection pocket will be provided. For counting, twelve unit counters will be provided, as follows:

- 10 four-position counters
- 1 six-position subtotal counter
- 1 six-position final total counter

Each counter will have flexible pluggable entry and reset control. Any detected expression or combination of expressions may be counted, irrespective of the basis of card selection. It will be possible, therefore, to count on bases broader or narrower than the basis of card selection, or on bases entirely different from the basis of card selection, in order to obtain exploratory information.

Security Information

f. Technique for Presenting Search Requirements to Machine

Search requirements will be presented to the searching machine by way of a punched search card in conjunction with plugboard manipulations.

1. Search Card

When the punching code employed for recording is one of the basic alphabets (i.e., standard IBM code or IBM scanning code), the basic patterns required for a search will be presented to the machine in the form of a punched search card or "question card." Conventional IBM cards will be used for this purpose.

(a) When standard IBM code is used for recording, the final search card will be punched in an auxiliary code appropriate for establishing matching circuits. This card will be prepared automatically by means of a preliminary search card punched in standard code and a standard duplicating key punch, slightly modified to provide for automatic conversion to the auxiliary code.

(b) When IBM scanning code is used for recording, an auxiliary punching code for searching will not be required.

2. Plugboard Manipulations

Search requirements other than required patterns from a basic alphabet will be presented to the machine by way of a plugboard. The requirements so handled will include the following:

(a) organization of patterns into expressions of more than one pattern each

(b) organization of expressions into combinations of expressions

(c) specification of boundary conditions

(d) specification of output functions (selection, counting)

(e) specification of special (non-basic) patterns

Security Information

g. Provisions for Searching Sources Represented by
More Than One Card

As indicated in section a3, the number of record cards punched for a single source may vary from one card to any needed number of cards.

1. Means will be provided whereby the scanner will recognize the beginning and end of a record, regardless of the number of cards involved.
2. Checking mechanisms will be provided which may be set to issue a "stop machine" signal if a set of cards representing a single source is

(a) incomplete, or

(b) out of proper internal order.

Use of these checking mechanisms will be optional.

3. A "transit storage" device will be incorporated in the card feeding unit so that a set of cards representing a single source may be selected as a set, without disruption or derangement. This device will have a holding capacity of twenty-five (25) cards and will function as follows:

(a) If a signal to select occurs prior to the twenty-sixth card of a set, all cards of that set will be selected automatically, regardless of the total number.

(b) If a signal to select occurs after the twenty-fifth card of a set, the group of cards selected automatically will comprise, in proper order, card x, and the twenty-four cards directly preceding card x. Simultaneously, a "stop machine" signal may be generated, if desired, to indicate to the machine operator that a set of cards has been split.

h. Provision for Searching Conventional Files of IBM Cards

The searching machine will be equipped with a means of detecting sought expressions in specified columnar or field locations. The purpose of this provision is to enable the searching of record cards which have been coded and punched on a fixed-field basis for processing in conventional IBM equipment. As a consequence of this provision, it will not be necessary to re-code, re-punch or modify conventional files in order to take advantage of the speed, capacity, and flexibility of the scanner in searching operations.

PUNCHING CODES: Standard IBM Code

DIGITS										LETTERS																										SPECIAL CHARACTERS					
0123456789										ABCDEFGHIJKLMN O PQRSTU VWXYZ																										& . - \$ % / , # @					
0000000000										00000000000000000000000000000000																															
1111111111										11111111111111111111111111111111																															
2222222222										22222222222222222222222222222222																															
3333333333										33333333333333333333333333333333																															
4444444444										44444444444444444444444444444444																															
5555555555										55555555555555555555555555555555																															
6666666666										66666666666666666666666666666666																															
7777777777										77777777777777777777777777777777																															
8888888888										88888888888888888888888888888888																															
9999999999										99999999999999999999999999999999																															
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0		

FIGURE 1